wwPDB X-ray Structure Validation Summary Report (i)

Apr 11, 2023 – 02:35 pm BST

PDB ID	:	8COB
Title	:	Crystal structure of human PCNA in complex with ERCC6L2 PIP box peptide
Authors	:	Ariza, A.
Deposited on	:	2023-02-27
Resolution	:	2.73 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

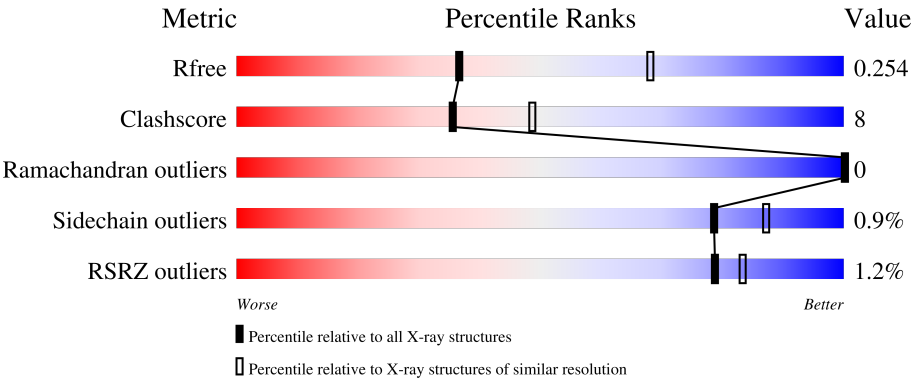
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.32.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.32.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.73 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1271 (2.76-2.72)
Clashscore	141614	1322 (2.76-2.72)
Ramachandran outliers	138981	1297 (2.76-2.72)
Sidechain outliers	138945	1298 (2.76-2.72)
RSRZ outliers	127900	1243 (2.76-2.72)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	261	76% 18% 5%
1	C	261	3% 72% 23% • 5%
1	E	261	79% 16% 5%
2	B	15	67% 7% 27%
2	D	15	60% 13% 27%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	15	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 6252 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Proliferating cell nuclear antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	247	Total	C	N	O	S	0	11	0
			1981	1247	325	393	16			
1	C	249	Total	C	N	O	S	0	10	0
			1982	1247	322	397	16			
1	E	249	Total	C	N	O	S	0	10	0
			1978	1244	321	397	16			

- Molecule 2 is a protein called DNA excision repair protein ERCC-6-like 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	11	Total	C	N	O	S	0	0	0
			80	51	13	15	1			
2	D	11	Total	C	N	O	S	0	0	0
			80	51	13	15	1			
2	F	12	Total	C	N	O	S	0	0	0
			87	56	14	16	1			

- Molecule 3 is water.

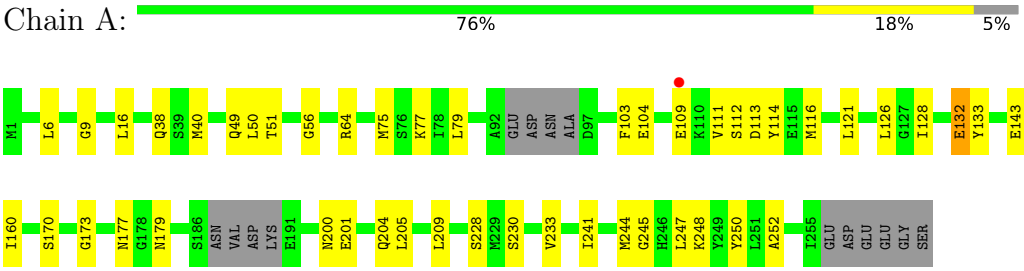
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	23	Total	O	0	0
			23	23		
3	C	14	Total	O	0	0
			14	14		
3	E	24	Total	O	0	0
			24	24		
3	F	3	Total	O	0	0
			3	3		

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

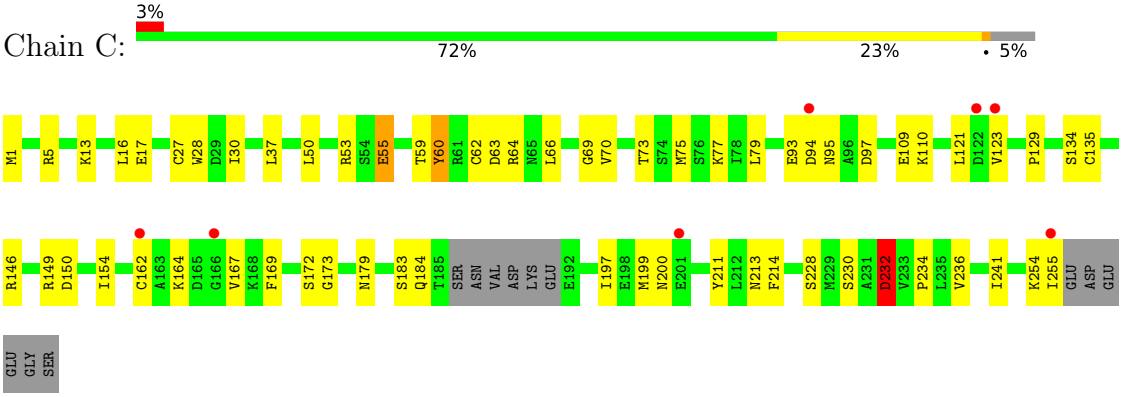
• Molecule 1: Proliferating cell nuclear antigen

Chain A:



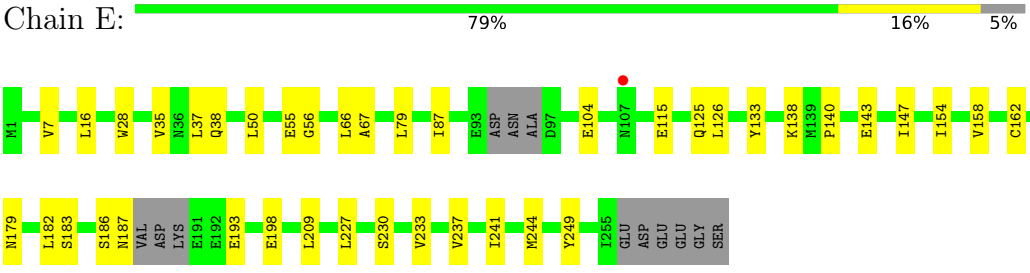
• Molecule 1: Proliferating cell nuclear antigen

Chain C:



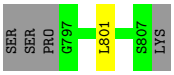
• Molecule 1: Proliferating cell nuclear antigen

Chain E:



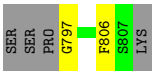
• Molecule 2: DNA excision repair protein ERCC-6-like 2

Chain B:



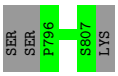
• Molecule 2: DNA excision repair protein ERCC-6-like 2

Chain D:



• Molecule 2: DNA excision repair protein ERCC-6-like 2

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	67.69Å 109.08Å 139.40Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.56 – 2.73 48.56 – 2.73	Depositor EDS
% Data completeness (in resolution range)	99.8 (48.56-2.73) 97.9 (48.56-2.73)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.82 (at 2.73Å)	Xtriage
Refinement program	PHENIX dev_3126	Depositor
R, R_{free}	0.197 , 0.255 0.197 , 0.254	Depositor DCC
R_{free} test set	1405 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	46.3	Xtriage
Anisotropy	0.264	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 52.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6252	wwPDB-VP
Average B, all atoms (Å^2)	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.39% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$\|Z\|$ >5	RMSZ	#$\|Z\|$ >5
1	A	0.37	1/2024 (0.0%)	0.66	2/2728 (0.1%)
1	C	0.46	2/2026 (0.1%)	0.79	5/2735 (0.2%)
1	E	0.36	0/2021	0.60	0/2726
2	B	0.30	0/80	0.63	0/106
2	D	0.31	0/80	0.56	0/106
2	F	0.38	0/88	0.51	0/117
All	All	0.40	3/6319 (0.0%)	0.68	7/8518 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	2
1	C	0	3
1	E	0	1
All	All	0	6

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	55	GLU	CB-CG	9.15	1.69	1.52
1	A	77	LYS	CE-NZ	7.35	1.67	1.49
1	C	135	CYS	CB-SG	5.07	1.90	1.82

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	232	ASP	CB-CG-OD2	-16.16	103.76	118.30
1	C	232	ASP	CB-CG-OD1	10.92	128.12	118.30
1	C	55	GLU	N-CA-CB	-6.49	98.91	110.60
1	A	77	LYS	CD-CE-NZ	-6.47	96.83	111.70

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	121	LEU	CA-CB-CG	5.58	128.14	115.30

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	132	GLU	Peptide
1	A	241	ILE	Peptide
1	C	232	ASP	Sidechain
1	C	241	ILE	Peptide
1	C	95	ASN	Peptide

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1981	0	2000	35	0
1	C	1982	0	1992	43	0
1	E	1978	0	1989	29	0
2	B	80	0	80	1	0
2	D	80	0	80	6	0
2	F	87	0	88	0	0
3	A	23	0	0	0	0
3	C	14	0	0	0	0
3	E	24	0	0	0	0
3	F	3	0	0	0	0
All	All	6252	0	6229	100	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 8.

The worst 5 of 100 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:53:ARG:HD2	1:C:55:GLU:OE1	1.69	0.92

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:E:56:GLY:HA3	1:E:244:MET:HG2	1.66	0.78
1:A:6:LEU:HD12	1:A:9:GLY:HA2	1.64	0.78
1:C:53:ARG:HB3	1:C:55:GLU:HG3	1.70	0.74
1:E:140:PRO:HG3	1:E:193:GLU:HA	1.68	0.74

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	252/261 (97%)	251 (100%)	1 (0%)	0	100	100
1	C	255/261 (98%)	247 (97%)	8 (3%)	0	100	100
1	E	253/261 (97%)	250 (99%)	3 (1%)	0	100	100
2	B	9/15 (60%)	9 (100%)	0	0	100	100
2	D	9/15 (60%)	9 (100%)	0	0	100	100
2	F	10/15 (67%)	10 (100%)	0	0	100	100
All	All	788/828 (95%)	776 (98%)	12 (2%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	227/228 (100%)	224 (99%)	3 (1%)	69 82
1	C	227/228 (100%)	224 (99%)	3 (1%)	69 82
1	E	228/228 (100%)	227 (100%)	1 (0%)	91 94
2	B	9/13 (69%)	9 (100%)	0	100 100
2	D	9/13 (69%)	9 (100%)	0	100 100
2	F	10/13 (77%)	10 (100%)	0	100 100
All	All	710/723 (98%)	703 (99%)	7 (1%)	78 85

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	60	TYR
1	C	172	SER
1	E	162	CYS
1	C	232	ASP
1	A	143[B]	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
2	F	798	GLN
1	E	187	ASN
1	C	184	GLN
1	A	184	GLN
1	C	200	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95^{th} percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	A	247/261 (94%)	-0.04	1 (0%) 92 95	31, 61, 109, 131	0
1	C	249/261 (95%)	0.00	7 (2%) 53 60	35, 61, 115, 153	0
1	E	249/261 (95%)	-0.17	1 (0%) 92 95	24, 51, 94, 140	0
2	B	11/15 (73%)	-0.08	0 100 100	49, 58, 66, 75	0
2	D	11/15 (73%)	0.09	0 100 100	76, 84, 96, 102	0
2	F	12/15 (80%)	-0.41	0 100 100	29, 43, 63, 98	0
All	All	779/828 (94%)	-0.07	9 (1%) 79 83	24, 58, 109, 153	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	123	VAL	3.2
1	C	201	GLU	3.2
1	C	122	ASP	3.1
1	C	94	ASP	2.9
1	C	255	ILE	2.5

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.